Efficient Importance Sampling for Rare Event Simulation with Applications

Cheng-Der Fuh[*],   Huei-Wen Teng[†]   and   Ren-Her Wang[‡]

February 5, 2013

Abstract

Importance sampling has been known as a powerful tool to reduce the variance of Monte Carlo estimator for rare event simulation. Based on the criterion of minimizing the variance of Monte Carlo estimator within a parametric family, we propose a general account for finding the optimal tilting measure. To this end, when the moment generating function of the underlying distribution exists, we obtain a simple and explicit expression of the optimal alternative distribution. The proposed algorithm is quite general to cover many interesting examples, such as normal distribution, noncentral χ^2 distribution, and compound Poisson processes. To illustrate the broad applicability of our method, we study value-at-risk (VaR) computation in financial risk management and bootstrap confidence regions in statistical inferences.

Some key words: Bootstrap; confidence region; exponential tilting; moderate deviation; VaR.

[*]Chair Professor at the Graduate Institute of Statistics, National Central University, No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan (R.O.C.). Email:cdfuh@ncu.edu.tw

[†]Assistant Professor at the Graduate Institute of Statistics, National Central University, No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan (R.O.C.). Email:wenteng@ncu.edu.tw

[‡]Assistant Professor at the Department of Banking and Finance, Tamkang University, New Taipei City, 25137 Taiwan (R.O.C.). Email:138230@mail.tku.edu.tw

1   Introduction

This paper considers the problem of estimating small probabilities by Monte Carlo simulations. That is, we estimate $z = P(A)$ when z is small, say of the order 10^{-2} or 10^{-3} or so; i.e., A is a moderate deviation rare event. Such problems appear in the construction of confidence regions for asymptotically normal statistics; cf. Beran (1987), Beran and Millar (1986), Hall (1987, 1992), Fuh and Hu (2004), and in the computation of value-at-risk (VaR) in risk management; cf. Jorion (2001), Duffie and Singleton (2003), Glasserman et al. (2000, 2002), Fuh et al. (2011). It is well known that importance sampling, where one uses observations from an alternative distribution Q to estimate the target distribution P, is a powerful tool in efficient simulation of events with small probabilities. Some general references are in Heidelberger (1995), Liu (2001), and Asmussen and Glynn (2007).

A useful tool in importance sampling for rare event simulation is exponential tilting, cf. Siegmund (1976), and Bucklew (2004) and references therein. The above mentioned algorithm is more efficient for large deviation rare event, i.e., z is of the order 10^{-5} or less. Examples of such events occur in telecommunications (z = bit-loss rate, probability of buffer overflow) and reliability (z = the probability of failure before time t). To be more precise, when a sequence of random vectors $\{X_n\}$ converge to a constant vector μ, for any event A not containing μ, the probability $P\{X_n \in A\}$ usually decays exponentially fast as $n \to \infty$. Efficient Monte Carlo simulation of such events has been obtained by Sadowsky and Bucklew (1990) based on the large deviations theory given by Ney (1983).

For moderate deviation rare event simulations, efficient importance sampling has been studied by Johns (1988), Davison (1988), Do and Hall (1991), and Fuh and Hu (2004, 2007). However, those papers concern one- and/or multivariate-normal distributions. Extension to heavy-tailed settings such as multivariate t distribution can be found in Fuh et al. (2011). The goal of this paper is to provide a simple general account for exponential tilting importance sampling, which covers all previous results and many other interesting examples.

It is worth mentioning that for events of large deviations $P\{X \in A\}$, Sadowsky and Bucklew (1990) showed that the asymptotically optimal alternative distribution is obtained through exponential tilting; that is, $Q(dx) = C \exp(\theta x) P(dx)$, where C is a normalizing constant and θ determines the amount of tilting. The optimal amount of tilting θ is such that the expectation of X under Q-measure equals the dominating point located at the boundary of A. However, for moderate deviation rare event, we show that typically the tilting point of the optimal alternative distribution is in the interior of A, which is different from the dominating point of large deviations theory. Furthermore, by using the idea of conjugate measure of Q, $\bar{Q}(dx) = C \exp(-\theta x) P(dx)$, the general account of our approach characterizes the optimal tilting θ, by solving the equation of the expectation of X under Q-measure equals the conditional expectation of X under $\bar{Q}$-measure given the rare event.

There are three aspects in this study. To begin with, we obtain an explicit expression for the optimal alternative distribution under exponential embedding family. Second, the proposed algorithm is quite general to cover many interesting examples, such as normal distribution, non-central χ^2 distribution, and compound Poisson processes. Third, the derived tilting formula can be used to calculate portfolio VaR under jump diffusion models, and to approximate bootstrap

confidence regions of parameters in regression models.

The rest of this paper is organized as follows. In Section 2, we present a general account of importance sampling that minimizes the variance of the Monte Carlo estimator within a parametric family, provide a recursive algorithm for finding the optimal alternative distribution, and approximate optimal tilting probability measure for moderate deviation events. Section 3 presents several frequently used examples to which we can characterize the optimal titling probability measures, and reports the relative efficiency of the proposed method with respect to the naive Monte Carlo through a simulation study. In Section 4, we demonstrate the performance of the tilting formula by investigating two examples: calculating portfolio VaR and bootstrapping confidence regions. Concluding remarks are given in Section 5. The proofs are deferred to the appendix.

2 Importance Sampling

2.1 A general account in importance sampling

Let $(\Omega, \mathcal{F}, P)$ be a given probability space, X be a random variable on Ω and A be a measurable set in $\mathbf{R}$. To estimate the probability of an event $\{X \in A\}$, we shall employ the importance sampling method. That is, instead of sampling from the target distribution P of X directly, we sample from an alternative distribution $Q := Q_\theta$. Suppose X has moment generating function $\Psi(\theta) = E[e^{\theta X}]$ under P for $\theta \in \mathbf{R}$. Then we consider the exponential tilting measure Q of P, which has the form

$$\frac{dQ}{dP} = \frac{e^{\theta X}}{E[e^{\theta X}]} = e^{\theta X - \psi(\theta)},$$

where $\psi(\theta)$ is $\log \Psi(\theta)$, the cumulant generating function of X. The question is how to choose an alternative distribution Q so that the importance sampling estimator has the minimum variance.

The importance sampling estimator for $p = P\{X \in A\}$ based on a sample of size n is

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \in A\}} \frac{dP}{dQ}, \tag{2.1}$$

where $\mathbf{1}_B$ is the indicator function of an event B, X_i, $i = 1, \ldots, n$, are independent observations from Q, and dP/dQ is the Radon-Nikodym derivative assuming P is absolutely continuous with respect to Q. Set

$$G(\theta) = E_Q\left[\mathbf{1}_{\{X \in A\}} \frac{dP}{dQ}\right]^2 = E\left[\mathbf{1}_{\{X \in A\}} \frac{dP}{dQ}\right] = E[\mathbf{1}_{\{X \in A\}} e^{-\theta X + \psi(\theta)}],$$

where the expectation without any qualification is under the target probability measure P unless otherwise stated. Observing that, since the estimator $\hat{p}_n$ is unbiased, the variance of the importance sampling estimator is

$$\text{var}_Q(\hat{p}_n) = n^{-1}(G(\theta) - p^2).$$

3

We remark that when $A = \{\omega : X(\omega) > a\}$ for some constant $a > 0$, large deviation theory considers the following inequality

$$G(\theta) \le e^{-\theta a + \psi(\theta)},$$

and minimizes the above upper bound. The first-order condition (after taking logarithm) gives

$$\psi'(\theta) = a. \tag{2.2}$$

Note that the approximation (2.2) is more accurate when $P\{X > a\}$ is small, i.e., a is sufficient large. In contrast to the large deviation theory, our goal is to solve the optimization problem for θ

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, G(\theta). \tag{2.3}$$

θ^* is hence the desired quantity for selecting the tilting measure that minimizes the variance of the importance sampling estimator within a suitable parametric family.

To minimize $G(\theta)$, the first-order condition gives

$$\frac{\partial G(\theta)}{\partial \theta} = E[\mathbf{1}_{\{X \in A\}} e^{-\theta X + \psi(\theta)}(-X + \psi'(\theta))] = 0.$$

Dividing $e^{\psi(\theta)}$ for both sides of the above equation, we have an equivalent condition,

$$E[\mathbf{1}_{\{X \in A\}} e^{-\theta X}(-X + \psi'(\theta))] = 0, \tag{2.4}$$

and therefore θ^* is the solution of

$$\psi'(\theta) = \frac{E[\mathbf{1}_{\{X \in A\}} X e^{-\theta X}]}{E[\mathbf{1}_{\{X \in A\}} e^{-\theta X}]}. \tag{2.5}$$

Ideally, closed-formed formulas of the right-hand-side (RHS) of (2.5) needs to be derived via analytic procedures. Then, standard numerical procedures (or a recursive algorithm presented in Section 2.2) can be applied to find θ^* satisfying (2.5). However, the derivation for a closed-formed formula of the RHS of (2.5) may be tedious and complicated in general cases. In this respect, the tilting measure given in (2.2) in the large deviation theory seems to be preferred because of its simplicity. To prove that the RHS of (2.5) can be represented in a simple formula, we consider the conjugate measure $\bar{Q} := \bar{Q}_\theta$ of the measure Q, which is defined as

$$\frac{d\bar{Q}}{dP} = \frac{e^{-\theta X}}{E[e^{-\theta X}]} = e^{-\theta X - \tilde{\psi}(\theta)}, \tag{2.6}$$

where $\tilde{\psi}(\theta)$ is $\log \tilde{\Psi}(\theta)$ with $\tilde{\Psi}(\theta) = E[e^{-\theta X}]$.

To present a connection between $\bar{Q}$ and Q, we consider their probability densities with respect to Lebesgue measure $\mathcal{L}$. It is straightforward to see that $\tilde{\psi}(\theta) = \psi(-\theta)$, which implies

$$\frac{d\bar{Q}_\theta}{d\mathcal{L}} = e^{-\theta x - \tilde{\psi}(\theta)} \frac{dP}{d\mathcal{L}} = e^{-\theta x - \psi(-\theta)} \frac{dP}{d\mathcal{L}} = e^{(-\theta)x - \psi(-\theta)} \frac{dP}{d\mathcal{L}} = \frac{dQ_{-\theta}}{d\mathcal{L}}. \tag{2.7}$$

4

The following theorem states the existence and uniqueness for the optimization procedure (2.3), and provides a simplification on the RHS of (2.5) using $\bar{Q}$. Before that, we need a condition for $\Psi(\theta)$ being steep to ensure the finiteness of the moment generating function $\Psi(\theta)$. To define steepness, let $\theta_{\max} := \sup\{\theta : \Psi(\theta) < \infty\}$ (for light-tailed distributions, we have $0 < \theta_{\max} \leq \infty$). Then steepness means $\Psi(\theta) \to \infty$ as $\theta \to \theta_{\max}$.

Theorem 1 *Suppose the moment generating function $\Psi(\theta)$ of X exists and second order continuously differentiable for $\theta \in \mathbf{I} \subset \mathbf{R}$. Furthermore, assume that $\Psi(\theta)$ is steep and $E[X|X \in A] > E(X) := \mu$. Then there exists a unique solution for the optimization problem (2.3), which satisfies*

$$\psi'(\theta) = E_{\bar{Q}_\theta}[X|X \in A]. \tag{2.8}$$

The proof of Theorem 1 will be given in the Appendix.

In summary, the three-step procedures to find θ^* are

1. Calculate the cumulant generating function $\psi(\theta)$ of X and its derivative $\psi'(\theta)$.

2. Find the exponential tilting measure Q.

3. Find θ^* as the solution of $\psi'(\theta) = E_{\bar{Q}_\theta}[X|X \in A]$.

Remark 1: Note that the optimal tilting point θ^* obtained in (2.8) highlights the fact that the tilting probability measure depends on the likelihood ratio (or the embedding probability Q in terms of $\bar{Q}$), the region, and the statistics of interest. Simple cases such as normal distribution and t-distribution are in Fuh and Hu (2004), and Fuh et al. (2011), respectively, by using the technique of change of variables. The idea of using conjugate measure in Theorem 1 seems to be new and simple according to our best knowledge.

Remark 2: In addition, characterization (2.8) is easy to implement as will be illustrated in Sections 3 and 4. Furthermore, it provides an insightful interpretation of θ^*, which can be used to compare with the large deviation tilting. To be more specific, when $A = \{w : X(w) > a\}$ for $a > \mu$, it is known that θ^* satisfies $\psi'(\theta) = E_{Q_\theta} X$. (2.8) indicates that the optimal tilting θ^* satisfies

$$E_{Q_\theta} X = E_{\bar{Q}_\theta}[X|X > a] > a. \tag{2.9}$$

A comparison between (2.9) with (2.2) shows that the large deviation tilting parameter is the dominating point a; while the optimal tilting θ^* is inside the region of $\{w : X(w) > a\}$. Similar interpretation can be applied to the case of A is a convex set in $\mathbf{R}^q$. However, when $A = \{w : X(w) \in (-a, a)^c\}$, where c denotes the complement, it is nature to spit $A = \{w : X(w) > a\} \cup \{w : X(w) < -a\}$ and apply importance sampling for each part. A simple guidance for the stratification, based on relative efficiency, can be found in Fuh and Hu (2004).

Remark 3: To approximate $z = EZ$ where $Z \geq 0$ P-a.s.. Define $P^*(dw) = \frac{|Z|}{E|Z|}P(dw)$ and $L^* = \frac{E|Z|}{|Z|}$. It is known (cf. Theorem 1.2 in Chapter V, Asmussen and Glynn, 2007) that the importance sampling estimator ZL^* under P^* has zero variance. By using similar idea, tilting probability obtained by (2.8) can be regarded as an approximation of the zero variance importance sampling estimator within a parametric class.

2.2 Calculating the optimal alternative distribution

Before employing the importance sampling method, it is first necessary to identify the optimal alternative distribution. Since the optimal θ in (2.8) cannot be computed directly, to find θ in the third step, we consider a simple recursive procedure for a general equation,

$$g(\theta) = h(\theta), \tag{2.10}$$

for some functions $g(\theta)$ and $h(\theta)$. In our setting, $g(\theta) = \psi'(\theta)$ and $h(\theta) = E_{\bar{Q}}[X|X \in A]$.

A simple recursive procedure is implemented as follows.

1. Start with an arbitrary θ^0. Set $i = 1$.

2. Calculate $t^i = g(\theta^{i-1})$ and find θ^i as the solution of $h(\theta) = t^i$.

3. Set $i = i + 1$. Return to 2 until $(\theta^{i+1} - \theta^i)/\theta^i$ becomes very small.

Because the exact objective function $G(\cdot)$ is difficult to optimize directly, the proposed method replaces G with a quasi objective function and does optimization on the latter. In the preceding algorithm, the initial value $\theta^{(0)}$ can be chosen as a *dominating* point of the event $\{X \in A\}$. Furthermore, if $\theta^{(0)}$ is sufficiently large, the density of X decreases rapidly and the solution of (2.5) is close to $\theta^{(0)}$. Therefore, fast convergence of the recursive algorithm is to be expected.

Let θ^* be the solution to (2.10), i.e., $g(\theta^*) = h(\theta^*)$. By using an argument similar to Theorem 2 of Fuh et al. (2011), we have the following proposition.

Proposition 1 *Choose a dominating point of the event $\{X \in A\}$ as the initial value $\theta^{(0)}$. Then*
  i) *the recursive algorithm either converges to θ^* or alternates in the limit between a pair of values $\underline{\theta} \neq \bar{\theta}$ satisfying*
$$g(\underline{\theta}) = h(\bar{\theta}) \quad \text{and} \quad h(\underline{\theta}) = g(\bar{\theta}). \tag{2.11}$$
  ii) *If there does not exist $\underline{\theta} \neq \bar{\theta}$ such that (2.11) holds, then the recursive algorithm converges to the solution of (2.10).*

In this subsection, we introduce a simple recursive algorithm. Alternative root-finding algorithms such as bisection method, Newton's method, secant method, etc. can also be used to find the root in (2.10).

2.3 Approximating optimal tilting probability measure

In this subsection, we get the optimal tilting probability measure by approximating $\psi'(\theta)$ for a moderate deviation event. Let $X_1, \ldots, X_n$ be i.i.d. random variables from a distribution function F, with mean μ and variance σ^2. We want to estimate by simulation of the probability

$$P(S_n/n \leq a_n), \tag{2.12}$$

for some a_n, where $S_n = \sum_{i=1}^{n} X_i$. Note that the probability is small when $a_n - \mu < 0$, and $a_n - \mu = O(1)$. Assume that the moment generating function of X_1 exists for some θ belonging

to some interval Θ which contains the origin, and let $\psi(\theta) := \log E(\exp(\theta X_1))$ denote the cumulant generating function of X_1. Note that $\mu = \psi'(0)$, $\sigma^2 = \psi''(0)$. By using the technique developed in Section 2.1, we first embed the original probability P in the following exponential family

$$\frac{d\bar{Q}_\theta}{d\mathcal{L}}(x) = \exp(\theta x - \psi(\theta))\frac{dP}{d\mathcal{L}}(x). \tag{2.13}$$

The class of estimators considered for the probability defined in (2.12) is

$$\delta_n = I(S_n/n \le a_n)\prod_{j=1}^n \frac{dP(X_j)}{dQ_\theta(X_j)}, \tag{2.14}$$

where $\{X_j\}$ are distributed from Q_θ. By Theorem 1, we have that if $\mu = 0$ and $a_n = a < 0$, the optimal point θ^* in simulations is given by $\psi'(\theta^*) = E_{\bar{Q}(\theta^*)}[X|X > a]$. Or $\psi'(\theta^*) = a$ for large deviation tilting.

The starting point of the approximation is that $a_n - \mu = o(1)$, or that the probability (2.12) is not exceedingly small. For instance one may be interested in the probability $P(S_n - n\mu \le a\sigma\sqrt{n})$ for $a < 0$, and this leads to $a_n = \mu + a\sigma/\sqrt{n}$. The point θ^* can be approximated in the following manner by a one-term Taylor expansion of $\psi'(\theta^*)$:

$$\psi'(0) + \theta^*\psi''(0) \cong E_{\bar{Q}_{\theta^*}}[X|X > a_n] \implies \mu + \sigma^2\theta^* \cong E_{\bar{Q}_{\theta^*}}[X|X > a_n]. \tag{2.15}$$

The calculations above suggest that when $a_n - \mu \to 0$, expansions for (2.13) and the optimal choice of parameter θ^* can be obtained through the first few moments of F (more specifically, through the mean and variance) without appealing to the properties of moment generating function. Therefore it is illuminating to consider a slightly different approach to the problem of estimating (2.12), with $a_n = \mu + a\sigma/\sqrt{n}$, especially, for heavy-tailed distributions. An alternative approach for heavy-tailed distribution without moment generating function is via the method of transform random variables, cf. Asmussen and Glynn (2007).

Example: Pareto distribution. Let $f(x) = \alpha(1 + x)^{-\alpha-1}$ and $\tilde{f}(x) = \tilde{\alpha}(1 + x)^{-\tilde{\alpha}-1}$, where $\tilde{\alpha} \to 0$. Here we consider $\tilde{\alpha} = \alpha\theta(a)$, and $\theta(a)$ is a solution of

$$\mu + \sigma^2\theta = E_{\bar{Q}_\theta}[X|X > a_n]. \tag{2.16}$$

Remark 4: Note that in (2.13), we apply the idea of exponential embedding for a non-parametric distribution F. Further approximation can be done on (2.14) via the LAN family. This idea had been carried out in the bootstrap setting of Fuh and Hu (2004, 2007). More detailed analysis along this line and the comparison of non-paremetic importance sampling in Zhang (1996) and Neddermeyer (2009), for instance, will be published in a separate paper.

3 Examples and Simulation Study

3.1 Examples

To illustrate the general account of importance sampling, in this subsection, we study three examples: normal distribution, noncentral χ^2 distribution and compound Poisson processes,

and report several other interesting distributions. The simulation event is $\{X > a\}$ for a given random variable X and some $a > 0$. In these examples, we explicitly calculate a closed-form formula of $E_{\bar{Q}}[X | X \in A]$ when possible.

Example 1: Normal distribution

Let X be a random variable with standard normal distribution, denoted by $N(0, 1)$, with probability density function (pdf) $\frac{dP}{d\mathcal{L}} = e^{-x^2/2}/\sqrt{2\pi}$. Standard calculation gives $\Psi(\theta) = e^{\theta^2/2}$, $\psi(\theta) = \theta^2/2$ and $\psi'(\theta) = \theta$. In this case, the tilting measure Q is $N(\theta, 1)$, a location shift, and $\bar{Q}$ is $N(-\theta, 1)$. Applying Theorem 1 and using the fact that $X | \{X > a\}$ is a truncated normal distribution with minimum value a under $\bar{Q}$, θ^* needs to satisfy

$$\theta = \frac{\phi(a + \theta)}{1 - \Phi(a + \theta)} - \theta. \tag{3.1}$$

Alternatively, $G(\theta) = e^{\theta^2}(1 - \Phi(a + \theta))$, and θ^* must satisfy the first-order condition, $2\theta(1 - \Phi(a + \theta)) = \phi(a + \theta)$, or equivalently, $\theta = \phi(a + \theta)/2(1 - \Phi(a + \theta))$. By using $\frac{1 - \Phi(x)}{\phi(x)} \sim \frac{1}{x}$ as $x \to \infty$, it is easy to see from equation (3.1) that $\theta^* \sim a$ when a is large. This is the same as the large deviation tilting probability.

We remark the normal distribution has been analysed in Fuh and Hu (2004), for illustration, we consider this example from our general account and provide a simple and explicit characterization for θ^*, in the sense that the right-hand-side of (3.1) is a straightforward application of Theorem 1.

Example 2: Noncentral $\chi^2(\lambda, \kappa)$ distribution

Let Z_i be independent, normally distributed random variables with mean μ_i and variances σ_i^2, for $i = 1, \ldots, \kappa$. Then the random variable $X = \sum_{i=1}^{\kappa} \left(\frac{Z_i}{\sigma_i}\right)^2$ is distributed according to the noncentral χ^2 distribution. It has two parameters: κ which specifies the number of degrees of freedom, and λ, the noncentrality parameter, is defined as $\lambda = \sum_{i=1}^{\kappa} \left(\frac{\mu_i}{\sigma_i}\right)^2$.

The pdf of X is

$$\frac{dP}{d\mathcal{L}} = \frac{1}{2} e^{-(x+\lambda)/2} \left(\frac{x}{\lambda}\right)^{\kappa/4 - 1/2} I_{\kappa/2 - 1}(\sqrt{\lambda x}), \quad x \geq 0, \tag{3.2}$$

where $I_v(z)$ is a modified Bessel function of the first kind given by

$$I_v(z) = (z/2)^v \sum_{j=0}^{\infty} \frac{(z^2/4)^j}{j! \Gamma(v + j + 1)}.$$

Alternatively, the pdf of X can be written as

$$\frac{dP}{d\mathcal{L}} = \sum_{i=0}^{\infty} \frac{e^{-\lambda/2}(\lambda/2)^i}{i!} f_{Y_{\kappa + 2i}}(x), \quad x \geq 0, \tag{3.3}$$

where Y_q is a chi-square distribution with q degrees of freedom, denoted by $Y_q \sim \chi^2(q)$. By representation (3.3), the noncentral χ^2 distribution is a Poisson weighted mixture of central χ^2 distributions. Suppose that a random variable N has a Poisson distribution with mean $\lambda/2$, and the conditional distribution of X given $N = i$ is χ^2 with $\kappa + 2i$ degrees of freedom. Then the unconditional distribution of X is noncentral χ^2 with κ degrees of freedom, and noncentrality parameter λ, or,

$$X|\{N = i\} \sim \chi^2(\kappa + 2i), \quad N \sim Pois(\lambda/2).$$

Let the likelihood ratio $\frac{dQ}{dP} = e^{\theta X - \psi(\theta)}$, where $\psi(\theta) = \log E[\exp(\theta X)]$. To derive the tilting formula of X, note that $\Psi(\theta) = \exp\left(\frac{\lambda\theta}{1-2\theta}\right)/(1-2\theta)^{\kappa/2}$, $\psi(\theta) = \frac{\lambda\theta}{1-2\theta} - \frac{\kappa}{2}\log(1-2\theta)$, and $\psi'(\theta) = \frac{\lambda+\kappa(1-2\theta)}{(1-2\theta)^2}$. Therefore, the exponential tilting measure Q becomes

$$
\begin{aligned}
\frac{dQ}{d\mathcal{L}} &= \frac{dQ}{dP}\frac{dP}{d\mathcal{L}} = e^{\theta x - \psi(\theta)} \sum_{i=0}^{\infty} \frac{e^{-\frac{\lambda}{2}}(\frac{\lambda}{2})^i}{i!} \frac{1}{\Gamma(\frac{\kappa+2i}{2})2^{\frac{\kappa+2i}{2}}} x^{\frac{\kappa+2i}{2}-1}e^{-\frac{x}{2}} \qquad (3.4)\\
&= \sum_{i=0}^{\infty} \frac{e^{-\frac{\lambda}{2(1-2\theta)}}(\frac{\lambda}{2(1-2\theta)})^i}{i!} \frac{1}{\Gamma(\frac{\kappa+2i}{2})(\frac{2}{1-2\theta})^{\frac{\kappa+2i}{2}}} x^{\frac{\kappa+2i}{2}-1}e^{-\frac{x}{\frac{2}{1-2\theta}}}.
\end{aligned}
$$

Hence, by (3.4) X can be characterized by $X|\{N = i\} \sim \Gamma\left(\frac{\kappa+2i}{2}, \frac{2}{1-2\theta}\right)$ with $N \sim Pois\left(\frac{\lambda}{2(1-2\theta)}\right)$ under Q. By Theorem 1, θ^* is the solution of the following equation

$$\frac{\lambda + \kappa(1 - 2\theta)}{(1 - 2\theta)^2} = E_{\bar{Q}}[X|X > a], \qquad (3.5)$$

where X follows $X|\{N = i\} \sim \Gamma\left(\frac{\kappa+2i}{2}, \frac{2}{1+2\theta}\right)$ with $N \sim Pois\left(\frac{\lambda}{2(1+2\theta)}\right)$ under $\bar{Q}$.

When $\lambda = 0$, this reduces to $\chi^2(\kappa)$ distribution and the optimal θ^* needs to satisfy

$$\frac{\kappa}{1 - 2\theta} = E_{\bar{Q}}[X|X > a]. \qquad (3.6)$$

When the degree of freedom κ equals 2, $\chi^2(2)$ distribution reduces to the exponential distribution with mean 2. By the memoryless property, we have an explicit formula for $C(\theta)$ and hence θ^* needs to satisfy $\frac{2}{1-2\theta} = a + \frac{2}{1+\theta}$. As a result, we have an explicit solution for

$$\theta^* = \frac{-1 + \sqrt{1 + a^2}}{a} < 1. \qquad (3.7)$$

Note that the tilting formula (3.7) can also be found in Example 1 on p.22 of L'Ecuyer, Mandjes and Tuffin (2009).

Example 3: Compound Poisson process

Let $R_t = \sum_{n=1}^{N(t)} \log Y_n$ be the compound Poisson process, where jump event $N(t)$ is assumed to follow a Poisson process with parameter λ, and the jump sizes Y_n are assumed to follow a

lognormal distribution with parameters of location η and scale δ^2. Let $A = \{R_t > r_p\}$ for a given r_p, and compute

$$
\begin{aligned}
P(R_t > r_p) &= \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} P(\sum_{i=1}^{n} Z_i > r_p | N(t) = n) \\
&= \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} P(Z > \frac{r_p - n\eta}{\sqrt{n\delta^2}} | N(t) = n),
\end{aligned}
$$

where Z_i, $i = 1, \ldots, n$, are i.i.d. normal with mean η and variance δ^2 and $Z = \frac{\sum_{i=1}^{n} Z_i - n\eta}{\sqrt{n\delta^2}}$. Denote $f(R_t) = R_t - r_p$, and let the likelihood ratio

$$
\frac{dQ}{dP} = e^{\theta f(R_t) - \psi(\theta)}, \tag{3.8}
$$

where $\psi(\theta) = \log E[\exp(\theta f(R_t))]$.

Here the original measure P under $N(t) \sim Pois(\lambda t)$ and $\log Y_n \sim N(\eta, \delta^2)$ becomes

$$
\frac{dP}{d\mathcal{L}} = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^n e^{-\frac{\sum_{i=1}^{n}(z_i - \eta)^2}{2\delta^2}}.
$$

and

$$
\begin{aligned}
E[e^{\theta f(R_t)}] &= e^{-\theta r_p} E[e^{\theta \sum_{i=1}^{N(t)} Z_i}] = e^{-\theta r_p} \sum_{n=1}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} E[e^{\theta \sum_{i=1}^{n} Z_i} | N(t) = n] \\
&= e^{-\theta r_p} \sum_{n=1}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} e^{(\theta\eta + \frac{1}{2}\theta^2\delta^2)n} = e^{\lambda t(e^{\theta\eta + \frac{1}{2}\theta^2\delta^2} - 1) - \theta r_p}.
\end{aligned}
$$

Therefore, $\psi(\theta) = \lambda t(e^{\theta\eta + \frac{1}{2}\theta^2\delta^2} - 1) - \theta r_p$, and $\psi'(\theta) = \lambda t e^{\theta\eta + \frac{1}{2}\theta^2\delta^2}(\eta + \theta\delta^2) - r_p$. Hence, the likelihood ratio (3.8) becomes

$$
\frac{dQ}{dP} = e^{\theta \sum_{i=1}^{n} z_i} e^{-\lambda t(e^{\theta\eta + \frac{1}{2}\theta^2\delta^2} - 1)}.
$$

For a given $N(t) = n$,

$$
\begin{aligned}
\frac{dQ}{d\mathcal{L}} &= \frac{dQ}{dP} \frac{dP}{d\mathcal{L}} = e^{\theta \sum_{i=1}^{n} z_i} e^{-\lambda t(e^{\theta\eta + \frac{1}{2}\theta^2\delta^2} - 1)} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^n e^{-\frac{\sum_{i=1}^{n}(z_i - \eta)^2}{2\delta^2}} \\
&= \frac{(\lambda t e^{\theta\eta + \frac{1}{2}\theta^2\delta^2})^n e^{-\lambda t e^{\theta\eta + \frac{1}{2}\theta^2\delta^2}}}{n!} \left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^n e^{-\frac{\sum_{i=1}^{n}(z_i - (\eta + \theta\delta^2))^2}{2\delta^2}}.
\end{aligned}
$$

As a result, $R_t = \sum_{i=1}^{N(t)} Z_i$ with $Z_i \sim N(\eta + \theta\delta^2, \delta^2)$ and $N(t) \sim Pois(\lambda e^{\theta\eta + \theta^2\delta^2/2}t)$ under Q. By Theorem 1, θ^* needs to satisfy

$$
\lambda t e^{\theta\eta + \frac{1}{2}\theta^2\delta^2}(\eta + \theta\delta^2) - r_p = E_{\bar{Q}}[R_t | R_t > r_p],
$$

where $R_t = \sum_{i=1}^{N(t)} Z_i$ with $R_t | \{N(t) = n\} \sim N(\eta - \theta\delta^2, \delta^2)$ and $N(t) \sim Pois(\lambda e^{-\theta\eta + \theta^2\delta^2/2}t)$ under $\bar{Q}$.

Table 1 reports importance sampling tilting probability for eight distributions: Binomial, Poisson, exponential, normal, χ^2, Gamman, noncentral χ^2, and uniform. The case of t distribution can be found in Fuh et al. (2011). This table includes the family of tilting probability Q and $\psi'(\theta)$. Note that in Table 1, when $A = \{X > a\}$ for some $a > 0$, then $E_{\bar{Q}}[X|X > a] = a + \frac{1}{1+\theta}$ for exponential distribution $E(1)$ and $= \frac{\phi(a+\theta)}{1-\Phi(a+\theta)} - \theta$ for standard normal distribution. By Theorem 1, we have that $\bar{Q} = Q_{-\theta}$ in Examples 1–7.

Table 1: Summary of distributions and their tilting measures.

Ex	P	Q	$\psi'(\theta)$
1	$B(n,p)$	$B(n, \frac{pe^\theta}{pe^\theta + (1-p)})$	$\frac{npe^\theta}{1-p+pe^\theta}$
2	$Pois(\lambda)$	$Pois(\lambda e^\theta)$	λe^θ
3	$N(0, \sigma^2)$	$N(\theta, \sigma^2)$	$\theta\sigma^2$
4	$E(1)$	$E(\frac{1}{1-\theta})$	$\frac{1}{1-\theta}$
5	$\chi^2(\kappa)$	$\Gamma(\frac{\kappa}{2}, \frac{2}{1-2\theta})$	$\frac{\kappa}{1-2\theta}$
6	$\Gamma(\alpha, \beta)$	$\Gamma(\alpha, \frac{\beta}{1-\beta\theta})$	$\frac{\alpha\beta}{1-\beta\theta}$
7	$NC\chi^2(\kappa, \lambda)$	$X|\{N = i\} \sim \Gamma(\frac{\kappa+2i}{2}, \frac{2}{1-2\theta}), N \sim Pois(\frac{\lambda}{2(1-2\theta)})$	$\frac{\lambda+\kappa(1-2\theta)}{(1-2\theta)^2}$
8	$Uniform(0, 1)$	$\propto e^{\theta X} \mathbf{1}_{\{0 \leq X \leq 1\}}$	$\frac{e^\theta}{e^\theta - 1} - \frac{1}{\theta}$

Table 2: Relative efficiency for different distributions.

p	$N(0,1)$	$E(1)$	$\chi^2(1)$	$\Gamma(4,10)$	$\chi^2(2,10)$
0.0001	2409.74	818.53	603.61	1282.87	1529.46
0.001	290.90	109.88	82.74	166.00	192.20
0.01	38.06	16.57	12.90	23.74	26.54
0.05	9.98	4.99	4.04	6.76	7.34
0.1	5.77	3.13	2.60	4.10	4.38

3.2   Simulation Study

In this subsection, we present some numerical results on relative efficiency using the method outlined in Theorem 1 for estimating tail probabilities of $p = P\{X > a\}$ for some $a > 0$. The relative efficiency is defined as the variance ratio between the crude Monte Carlo estimator and the importance sampling estimator.

Table 2 compares the relative efficiency (RE) for various distributions. Here a is chosen such that $p = 0.1$ (central event), 0.05, 0.01 (95%, 99% confidence interval), 0.001 (VaR type probability) and 0.0001 (rare event). Monte Carlo sample size is $10,000$ for each case. Note that the relative efficiency in normal case is the largest.

Simulations are conducted for central $\chi^2(\kappa)$, noncentral $\chi^2(\lambda, \kappa)$ and gamma $\Gamma(\alpha, \beta)$ distributions. Since the results are almost the same, we only report the simulation results for noncentral $\chi^2(\lambda, \kappa)$ distribution in Table 3. We can see that the proposed method is significantly more efficient than the naive Monte Carlo in $10,000$ simulations for all probabilities of $p = 0.1$, 0.05, 0.01, 0.001 and 0.0001. Furthermore, the efficiency gain is larger for smaller probabilities against the naive method. The normal case can be found in Fuh and Hu (2004), which will be used, along with its square and compound Poisson model, for VaR computation. Normal distribution and χ^2 distribution will be used for importance resampling in bootstrapping confidence region of parameters in regression model.

Table 3: We compare the performance, in terms of means and standard errors (SE), of estimating the probability $P(NC\chi^2(\kappa, \lambda) > a)$ using the naive Monte Carlo method (MC) and our importance sampling method (IS) for a combination of parameters (κ, λ) and a variety of boundary values a with a sample size of 10,000. p is the theoretical value of the probability of interest. θ^* is the optimal tilting parameter. Both the relative efficiency (RE) and the benchmark RE (RE*) are reported, where the former is calculated as the ratio of the variance of the Monte Carlo estimate over that of the importance sampling estimate, and the later is calculated as $p(1-p)/(G(\theta^*) - p^2)$.

(κ, λ)	p	a	θ^*	MC		IS		RE	RE*
				mean	SE	mean	SE		
(2, 1)	0.0001	24.28	0.38	0.000000	0.000000	0.000097	0.000003	0.00	1059.49
	0.001	18.65	0.36	0.001200	0.000346	0.000981	0.000027	168.64	136.82
	0.01	12.85	0.33	0.010300	0.001010	0.009940	0.000226	20.04	19.51
	0.05	8.64	0.29	0.050100	0.002182	0.049579	0.000913	5.71	5.59
	0.1	6.77	0.26	0.101400	0.003019	0.099657	0.001610	3.52	3.42
(2, 5)	0.0001	37.05	0.31	0.000200	0.000141	0.000104	0.000003	2653.12	1381.16
	0.001	29.88	0.28	0.001200	0.000346	0.001045	0.000025	198.95	173.94
	0.01	22.23	0.25	0.009600	0.000975	0.009987	0.000202	23.28	24.22
	0.05	16.38	0.21	0.047500	0.002127	0.049684	0.000830	6.56	6.76
	0.1	13.64	0.19	0.098600	0.002981	0.099130	0.001493	3.99	4.07
(5, 1)	0.0001	30.02	0.36	0.000000	0.000000	0.000101	0.000003	0.00	1213.85
	0.001	24.07	0.34	0.001000	0.000316	0.001030	0.000026	152.60	156.69
	0.01	17.83	0.30	0.009800	0.000985	0.009918	0.000209	22.13	22.30
	0.05	13.17	0.26	0.049000	0.002159	0.049644	0.000856	6.36	6.37
	0.1	11.03	0.23	0.101400	0.003019	0.099369	0.001522	3.93	3.89
(5, 5)	0.0001	41.73	0.30	0.000100	0.000100	0.000101	0.000003	1401.51	1434.04
	0.001	34.35	0.28	0.000600	0.000245	0.001008	0.000024	106.09	181.10
	0.01	26.42	0.24	0.011800	0.001080	0.010086	0.000200	29.05	25.28
	0.05	20.29	0.20	0.046300	0.002101	0.048860	0.000805	6.81	7.05
	0.1	17.38	0.18	0.093900	0.002917	0.098820	0.001448	4.06	4.22

4   Applications

To illustrate the applicability of our proposed tilting formula, we present two applications in this section. In subsection 4.1, we apply the importance sampling for portfolio VaR computation, to which the tilting formula is given under the multivariate jump diffusion model of the underlying assets. In subsection 4.2, we study importance resampling for bootstrapping confidence regions in regression models. By using the idea of pseudo-maximum likelihood estimator (PMLE) and the criterion of minimizing variance of the Monte Carlo estimator under a parametric family, we propose a titling formula based on the parametric family of normal (or χ^2) distributions.

4.1   Evaluating Value at Risk

As a standard benchmark for market risk disclosure, VaR is the loss in market value over a specified time horizon that will not be exceeded with probability $1 - p$. Hence define VaR as the quantile l_p of the loss L in portfolio value during a holding period of a given time horizon Δt. To be more specific, we express the portfolio value $V(t, S(t))$ as a function of risk factors and time, where $S(t) = (S^{(1)}(t), \ldots, S^{(d)}(t))^T$ comprises the d risk factors to which the portfolio is exposed at time t and T denotes the transpose of a matrix. The loss of the portfolio over the time interval $[t, t + \Delta t]$ is $L = V(t, S(t)) - V(t + \Delta t, S(t + \Delta t))$. Therefore VaR, l_p, associated with a given probability p and time horizon Δt, is given by

$$P(L > l_p) = p. \tag{4.1}$$

Assume $S(t)$ follows a d-dimensional jump diffusion model such that the return processes are described by the stochastic differential equations

$$r_t^{(i)} = \frac{dS^{(i)}(t)}{S^{(i)}(t)} = \mu^{(i)}dt + \sigma^{(i)}dW^{(i)}(t) + \sum_{j=1}^{N(t)} \log Y_j^{(i)}, \ i = 1, 2, \ldots, d, \tag{4.2}$$

where $(W^{(1)}(t), \ldots, W^{(d)}(t))$ is a standard Brownian motion in R^d, $N(t) \sim Pois(\lambda dt)$ is a Poisson process, $\log Y_j^{(i)}$ are independent and identically distributed (i.i.d.) random variables with $N(\eta^{(i)}, \delta^{(i)^2})$ distribution. Here the drift parameters $\mu^{(i)}$, volatility parameters $\sigma^{(i)}$, jump frequency λ, and jump size parameters $\eta^{(i)}$, $\delta^{(i)}$ are given. Furthermore, we assume that the Brownian motion and the Poisson process are independent. Denote $\log Y_j \sim N(\underline{\eta}, \underline{\delta}^T \Sigma_J \underline{\delta})$, where

$$\underline{\eta} = \begin{bmatrix} \eta^{(1)} \\ \eta^{(2)} \\ \vdots \\ \eta^{(d)} \end{bmatrix}, \quad \underline{\delta} = \begin{bmatrix} \delta^{(1)} & 0 & \cdots & 0 \\ 0 & \delta^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \delta^{(d)} \end{bmatrix}, \Sigma_J = \begin{bmatrix} 1 & \rho_{12}^J & \cdots & \rho_{1d}^J \\ \rho_{12}^J & 1 & \cdots & \rho_{2d}^J \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1d}^J & \cdots & \cdots & 1 \end{bmatrix}.$$

For simple notations, denote

$$\underline{\mu} = \begin{bmatrix} \mu^{(1)}\Delta t \\ \vdots \\ \mu^{(d)}\Delta t \end{bmatrix}, \quad \underline{\sigma} = \begin{bmatrix} \sigma^{(1)}\sqrt{\Delta t} \\ \vdots \\ \sigma^{(d)}\sqrt{\Delta t} \end{bmatrix}, \quad X = \begin{bmatrix} X^{(1)} \\ \vdots \\ X^{(d)} \end{bmatrix} \sim MN\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \Sigma \right),$$

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1d} \\ \rho_{12} & 1 & \cdots & \rho_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1d} & \cdots & \cdots & 1 \end{bmatrix}, \quad J = \begin{bmatrix} \sum_{j=1}^{N(\Delta t)} \log Y_j^{(1)} \\ \sum_{j=1}^{N(\Delta t)} \log Y_j^{(2)} \\ \vdots \\ \sum_{j=1}^{N(\Delta t)} \log Y_j^{(d)} \end{bmatrix}.$$

A discrete version of (4.2) is

$$r_t^{(i)} = \mu^{(i)} \Delta t + \sigma^{(i)} \sqrt{\Delta t} X^{(i)} + \sum_{j=1}^{N(\Delta t)} \log Y_j^{(i)}, \quad i = 1, \ldots, d. \tag{4.3}$$

Next we shall describe a quadratic approximation to the loss L. Let $\underline{r} = (r^{(1)}, \ldots, r^{(d)})^T$ be a vector of return of d assets. Denote $\Delta S = [S(t + \Delta t) - S(t)] \approx \underline{r}$ as the change in S over the corresponding time interval. The delta-gamma methods refine the relationship between risk factors and portfolio value by including quadratic as well as linear terms. The delta-gamma approximation to the change in portfolio value is

$$V(t + \Delta t, S + \Delta S) - V(t, S) \approx \frac{\partial V}{\partial t} \Delta t + \delta^T \Delta S + \frac{1}{2} \Delta S^T \Gamma \Delta S,$$

where $\delta = (\delta_1, \ldots, \delta_d)^T$ and $\Gamma = [\Gamma_{ij}]_{i,j=1,\ldots,d}$, with

$$\delta_i = \frac{\partial V}{\partial S_i}, \quad \Gamma_{ij} = \frac{\partial^2 V}{\partial S_i \partial S_j}, \quad i, j = 1, \ldots, d,$$

and all derivatives are evaluated at the initial point (t, S). Hence we can approximate the loss

$$L \approx a_0 + a_1^T \underline{r} + \underline{r}^T A_1 \underline{r}, \tag{4.4}$$

where $a_0 = -\frac{\partial V}{\partial t} \Delta t$ is a scalar, $a_1 = -\delta$ is an d-vector and $A_1 = -\Gamma/2$ is a symmetric matrix.

Under the jump diffusion process assumption of $\underline{r}$, we have

$$\begin{aligned} L &= a_0 + a_1^T \underline{\mu} + a_1^T \underline{\sigma} X + a_1^T J + \underline{\mu}^T A_1 \underline{\mu} + \underline{\sigma}^T Z^T A_1 Z \underline{\sigma} \\ &= b_0 + a_1^T \underline{\sigma} X + a_1^T J + \underline{\sigma}^T Z^T A_1 X \underline{\sigma}, \end{aligned} \tag{4.5}$$

where $b_0 = a_0 + a_1^T \underline{\mu} + \underline{\mu}^T A_1 \underline{\mu}$ and $J = (\sum_{j=1}^{N(\Delta t)} \log Y_j^{(1)}, \ldots, \sum_{j=1}^{N(\Delta t)} \log Y_j^{(d)})^T$.

To have a simple approximation, we neglect the quadratic approximation of the jump part in (4.4) as it is very small compared to the return of portfolio. Let C be the square root of the positive definite matrix Σ such that $C^T C = \Sigma$. We can transform the distribution of X into Z, which is multivariate normal distributed with identity covariance matrix, so that $X = CZ$. Moreover, C can be chosen so that $C^T A_1 C$ is diagonalized with diagonal elements $\lambda_1, \ldots, \lambda_d$. Denote $L_b := L - b_0$, then

$$\begin{aligned} L_b &= a_1^T \underline{\sigma} X + a_1^T J + \underline{\sigma}^T X^T A_1 X \underline{\sigma} = a_1^T \underline{\sigma} CZ + \underline{\sigma}^T Z^T C^T A_1 CZ \underline{\sigma} + a_1^T J \\ &= b^T Z + \underline{\sigma}^T Z^T \Lambda Z \underline{\sigma} + a_1^T J = \sum_{j=1}^d b_j Z_j + \lambda_j (\sigma^{(j)})^2 \Delta t Z_j^2 + a_1^T J, \end{aligned} \tag{4.6}$$

15

where $b^T = a_1^T \underline{\sigma} C$.

By using the tilting formula developed in Theorem 1 of Section 2, we consider a family of alternative distributions. Let the likelihood ratio of the alternative probability measures with respect to the target measure P be of the form

$$\frac{dQ}{dP} = e^{\theta(L_b - r_p) - \psi(\theta)}, \tag{4.7}$$

where L_b is defined in (4.6), r_p is the pth quantile, and $\psi(\theta) = \log E[\exp(\theta(L_b - r_p))]$ is the cumulant generating function of $L_b - r_p$ under the target probability measure P. The domain of θ will be specified after Equation (4.12). The problem of finding the optimal alternative measure is then reduced to that of identifying the θ-value that yields the minimal variance for the importance sampling estimator.

We now proceed to find the probability density under the alternative measure Q. A simple calculation leads that

$$
\begin{aligned}
\psi(\theta) &= \log E(e^{\theta L_b - \theta r_p}) \\
&= \frac{1}{2} \sum_{i=1}^d \left(\frac{(\theta b_i)^2}{1 - 2\theta \lambda_i (\sigma^{(i)})^2 \Delta t} - \log(1 - 2\theta \lambda_i (\sigma^{(i)})^2 \Delta t) \right) \\
&\quad + \lambda \Delta t \left(e^{\theta a_1^T \underline{\eta} + \frac{1}{2}\theta^2 a_1^T (\underline{\delta}^T \Sigma_J \underline{\delta}) a_1} - 1 \right) - \theta r_p,
\end{aligned}
\tag{4.8}
$$

and

$$
\begin{aligned}
\psi'(\theta) &= \sum_{i=1}^d \left(\frac{\theta(b_i)^2 (1 - \theta \lambda_i (\sigma^{(i)})^2 \Delta t)}{(1 - 2\theta \lambda_i (\sigma^{(i)})^2 \Delta t)^2} + \frac{\lambda_i (\sigma^{(i)})^2 \Delta t}{1 - 2\theta \lambda_i (\sigma^{(i)})^2 \Delta t} \right) - r_p \\
&\quad + \lambda \Delta t e^{\theta a_1^T \underline{\eta} + \frac{1}{2}\theta^2 a_1^T (\underline{\delta}^T \Sigma_J \underline{\delta}) a_1} (a_1^T \underline{\eta} + \theta a_1^T (\underline{\delta}^T \Sigma_J \underline{\delta}) a_1).
\end{aligned}
\tag{4.9}
$$

From (4.7), and (4.8), it follows that the importance sampling is done with an exponential tilting measure. Therefore, given $N(\Delta t) = n$,

$$
\begin{aligned}
\frac{dQ}{d\mathcal{L}} &= \frac{dQ}{dP}\frac{dP}{d\mathcal{L}} = \prod_{j=1}^d \frac{1}{\sqrt{2\pi}\sigma_j(\theta)} \exp\left\{ -\frac{(z_j - \mu_j(\theta))^2}{2\sigma_j^2(\theta)} \right\} \times \frac{\left(\lambda(\theta)\right)^n e^{-\lambda(\theta)}}{n!} \\
&\quad \times \left(\frac{1}{(2\pi)^{d/2} |\underline{\delta}^T \Sigma_J \underline{\delta}|^{1/2}} \right)^n \exp\left\{ -\frac{1}{2}\sum_{i=1}^n (v(i) - \eta(\theta))^T (\underline{\delta}^T \Sigma_J \underline{\delta})^{-1} (v(i) - \eta(\theta)) \right\},
\end{aligned}
\tag{4.10}
$$

where

$$\mu_j(\theta) = \frac{\theta b_j}{1 - 2\theta \lambda_j (\sigma^{(i)})^2 \Delta t}, \quad \sigma_j^2(\theta) = \frac{1}{1 - 2\theta \lambda_j (\sigma^{(i)})^2 \Delta t}, \tag{4.11}$$

and

$$\lambda(\theta) = \lambda \Delta t e^{\theta a_1^T \underline{\eta} + \frac{1}{2}\theta^2 a_1^T (\underline{\delta}^T \Sigma_J \underline{\delta}) a_1}, \quad \eta(\theta) = \underline{\eta} + \theta a_1^T (\underline{\delta}^T \Sigma_J \underline{\delta}). \tag{4.12}$$

16

That is, in (4.10) Z_j, $j = 1, \ldots, d$, are independent $N(\mu_j(\theta), \sigma_j^2(\theta))$, and J_i are compound Poisson processes with jump frequency $\lambda(\theta)$ and jump size $V(i)$, $i = 1, \ldots, n$ has mean $\eta(\theta)$ and variance matrix $\underline{\delta}^T \Sigma_J \underline{\delta}$. To guarantee the rational of Q under exponential twisting of measures, the constant θ must satisfy $1 - 2\theta\lambda_{(1)}(\sigma^{(i)})^2 \Delta t > 0$ and $1 - 2\theta\lambda_{(d)}(\sigma^{(i)})^2 \Delta t > 0$, where $\lambda_{(1)} = \max_{1 \le i \le d} \lambda_i$ and $\lambda_{(d)} = \min_{1 \le i \le d} \lambda_i$.

To have an efficient importance sampling for approximating $P\{L_b > r_p\}$ for some $r_p > 0$, we need to characterize the optimal titling θ via Theorem 1. Before stating the result, we define some quantities that facilitate the presentation of it. In view of (4.11), define

$$\bar{\mu}_j(\theta) = \mu_j(-\theta), \quad \bar{\sigma}_j^2(\theta) = \sigma_j^2(-\theta), \quad \bar{\lambda}(\theta) = \lambda(-\theta), \quad \bar{\eta}(\theta) = \eta(-\theta). \tag{4.13}$$

Let $\bar{V}(j)$ has a d-variate normal distribution with mean $\bar{\eta}(\theta)$ and variance matrix $\underline{\delta}^T \Sigma_J \underline{\delta}$ and $N(\Delta t)$ follows a $Pois(\bar{\lambda}(\theta))$. Applying Theorem 1, it is straight forward to obtain the following corollary.

Corollary 1 *Let θ be such that $1 \pm 2\theta\lambda_{(1)}(\sigma^{(i)})^2 \Delta t > 0$ and $1 \pm 2\theta\lambda_{(d)}(\sigma^{(i)})^2 \Delta t > 0$. Under the quadratic approximation to a portfolio VaR, the optimal alternative distribution Q minimizing the variance of the importance sampling estimator has θ satisfying*

$$\psi'(\theta) = E_{\bar{Q}}[L_b | L_b > r_p], \tag{4.14}$$

where $\psi'(\theta)$ is in (4.9), $E_{\bar{Q}}$ is the expectation under P_θ, which has the form (4.10) but replaced by the three joint distributions: $N(\bar{\mu}_j(\theta), \bar{\sigma}_j^2(\theta))$, $Pois(\bar{\lambda}(\theta))$, and $\bar{V}(j)$.

The optimal θ_p satisfying (4.14) can be searched by standard numerical methods or the recursive algorithm presented in Section 2.2. Next we present the importance sampling algorithm for multi-variate jump diffusion model as follows:

1. Compute the θ_p such that $\psi'(\theta_p) = E_{\theta_p}[L_b | L_b > r_p]$.

2.   (i) Generate $(Z_1, \ldots, Z_d)^T \sim (N(\mu_1(\theta_p), \sigma_1^2(\theta_p)), \ldots, N(\mu_d(\theta_p), \sigma_d^2(\theta_p)))^T$.

    (ii) Generate $N(\Delta t) \sim Pois(\lambda(\theta_p))$.

   (iii) Given $N(\Delta t) = n$ generate

$$V(i) = (\log Y_i^{(1)}, \log Y_i^{(2)}, \ldots, \log Y_i^{(d)})^T \sim N(\eta(\theta_p), \underline{\delta}^T \Sigma_J \underline{\delta}), \quad i = 1, \ldots, n.$$

3. Repeat step 2 k times to have $L_{b,i}$ defined in (4.6) for $i = 1, \ldots, k$. Compute $\hat{p}(\theta_p) = \frac{1}{k} \sum_{i=1}^k 1_{\{L_{b,i} > r_p\}} e^{-\theta_p(L_{b,i} - r_p) + \psi(\theta_p)}$.

4. Repeat steps 2 and 3 with Monte Carlo size M, and compute sample variance $\widehat{var}_p(\theta_p)$.

In Table 4, we compare the relative efficiency (RE) with fifteen risk factors of the proposed method (PSD) with respect to the naive method (NV) in estimating the loss probabilities $P\{L_b > x\}$ with different values of x. Here the relative efficiency, RE(Method1, Method2), is the variance under Method 2 divided by that under Method 1. The parameters are $\lambda = 1$, $\eta^{(1)} = \eta^{(2)} = \ldots = \eta^{(15)} = 0$, $\delta^{(i)} = i/100$, $\mu^{(i)} = i/100$, $\sigma^{(i)} = 0.1 + i/100$ and $\rho_{ij} = 0.3$ for all $i \neq j$, sample size $k = 1,000$ and Monte Carlo replication $M = 10,000$. In Table 4, we set $b_1 = 0.044$, $b_2 = 0.0589891$, $b_3 = 0.0720326$, $b_4 = 0.0838734$, $b_5 = 0.0948957$, $b_6 = 0.105325$, $b_7 = 0.115304$, $b_8 = 0.124929$, $b_9 = 0.134271$, $b_{10} = 0.143378$, $b_{11} = 0.152289$, $b_{12} = 0.161034$, $b_{13} = 0.169635$, $b_{14} = 0.178112$, $b_{15} = 0.186479$, $\lambda_1 = 5.2$ and $\lambda_2 = \ldots = \lambda_{15} = 0.7$ with quadratic approximation function for a fifteen-variate jump diffusion model. We can see that the proposed method is significantly more efficient than the naive Monte Carlo in $10,000$ simulations for all values of x. Furthermore, the efficiency gain is larger for smaller probabilities against the naive method.

Table 4: Quadratic approximation function compared with naive method and a multi-variate jump diffusion model in Monte Carlo simulation.

r_p	0.824	1.166	1.549
p	0.0500	0.0100	0.00100
NV $\hat{p}$	0.0501	0.00998	0.00099
$\widehat{var}_p$	4.80E-05	1.00E-05	9.85E-07
PSD $\hat{p}(\theta)$	0.0501	0.00999	0.000100
$\widehat{var}_p(\theta)$	5.12E-06	2.69E-07	3.65E-09
RE$(\widehat{p}(\theta), \widehat{p})$	9.36	37.15	269.79

p denotes the true tail probability, r_p denotes the quantile of p, $\hat{p}$ and $\widehat{var}_p$ are the mean and variance of the probability estimator with Monte Carlo, $\hat{p}(\theta)$ and $\widehat{var}_p(\theta)$ are the mean and the variance of the tail probability estimator with importance sampling, RE$(\hat{p}(\theta), \hat{p})$ is the relative efficiency of $\hat{p}(\theta)$ relative to $\hat{p}$ in a multi-variate jump diffusion model.

4.2   Bootstrapping confidence regions in regression model

Consider a regression model

$$Y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i \quad \text{for} \quad i = 1, 2, \ldots, n, \tag{4.15}$$

where $p \geq 2$, $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. mean zero random variables with distribution $F(\cdot)$. Denote σ^2 as the variance of ε_1. Let $\mathbf{x_i} = (x_{i1}, \ldots, x_{ip})^T$, $\beta = (\beta_1, \ldots, \beta_p)^T$. In vector and matrix notation, we write $Y_i = \mathbf{x}_i^T \beta + \varepsilon_i$ for $i = 1, 2, \ldots, n$, and

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \tag{4.16}$$

where $\mathbf{X} = (x_{ij})_{n \times p}$, $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$. Then the least-squares estimator of β is

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{4.17}$$

Consider the problem of finding a confidence region C in $\mathbf{R}^p$ that covers the vector β with prescribed probability $(1 - \alpha)$. Denote β_0 as the true parameter. Under the assumption of $\mathbf{X}$ is full rank p, the statistics of interest is

$$T := \frac{(\hat{\beta} - \beta_0)^T(\mathbf{X}^T\mathbf{X})(\hat{\beta} - \beta_0)}{p\hat{\sigma}^2}, \tag{4.18}$$

where $\hat{\sigma}^2$ is the sample variance. In particular we would like to estimate the probability of the event $\{T \in A\}$, where A is chosen to be a circular confidence region for the unknown parameter β_0. Let $T^* = (\hat{\beta}^* - \hat{\beta})^T(\mathbf{X}^T\mathbf{X})(\hat{\beta}^* - \hat{\beta})/p\hat{\sigma}^{*2}$, where $\hat{\beta}^*$, $\hat{\sigma}^{*2}$ are the bootstrap estimators of $\hat{\beta}$ and $\hat{\sigma}^2$, respectively, for given data. Then the bootstrap estimator of $P\{T \in A\}$ is $\hat{u} = P\{T^* \in A | \hat{\varepsilon}\} = E(1_{\{T^* \in A\}} | \hat{\varepsilon})$, where $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)$.

To illustrate our proposed method, we select a sample from a regression model in Longley (1967) in which the data set is available at
http://www.itl.nist.gov/div898/strd/general/dataarchive.html. We choose this classical data set of labour statistics because it is one of the first used to test the accuracy of least squares computations. It is noted that the same algorithm can be applied to current statistical data set.

The response variable (y) is the Total Derived Employment and the predictor variables are GNP Implicit Price Deflator with Year 1954 = 100 (x_1), Gross National Product (x_2), Unemployment (x_3), Size of Armed Forces (x_4), Non-Institutional Population Age 14 & Over (x_5), and Year (x_6).

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \varepsilon_i \quad i = 1, 2, \ldots, n \tag{4.19}$$

where the ε_i is assumed by normal distribution with mean zero and variance σ^2 under PMLE. The least squares estimator (MLE) of β is

$$\hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_6) = (-3482258.63, 15.0618, -0.0358, -2.0202, -1.0332, -0.0511, 1829.15).$$

The sample variance $\hat{\sigma}^2$ is 55761.60. The bootstrap estimator of $P\{T > a\}$ is $\hat{\alpha} = P\{T^* > a | \hat{\varepsilon}\} = E(1_{\{T^* > a\}} | \hat{\varepsilon})$, where $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)$.

Let

$$T^* = \frac{(\hat{\beta}^* - \hat{\beta})^T(X^TX)(\hat{\beta}^* - \hat{\beta})}{p\hat{\sigma}^{*2}} = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^{*2}}{p\hat{\sigma}^{*2}},$$

where $\hat{\varepsilon}_i^*$ follows normal distribution with mean zero and variance $\hat{\sigma}^2$. Here $p = 6$ in equation (4.19).

To apply importance resampling in the setting, note that the original bootstrap distribution is

$$\frac{dP}{d\mathcal{L}} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\hat{\sigma}} e^{-\frac{\hat{\varepsilon}_i^{*2}}{2\hat{\sigma}^2}}.$$

The likelihood ratio based on an exponential change of measure is

$$\frac{dQ}{dP} = e^{\theta(T^* - a) - \psi_a(\theta)} = \frac{e^{\theta(T^* - a)}}{Ee^{\theta(T^* - a)}}, \qquad (4.20)$$

where $\psi_a(\theta) = \log Ee^{\theta(T^* - a)}$. Let $T_a = T^* - a$. The moment generating function of T_a is $E[e^{\theta T_a}] = e^{-\theta a}(1 - 2\frac{\theta}{p})^{-\frac{n}{2}}$. Therefore, the change of measure Q is

$$\frac{dQ}{d\mathcal{L}} = \frac{dQ}{dP}\frac{dP}{d\mathcal{L}} = e^{\theta T^*}(1 - 2\frac{\theta}{p})^{-\frac{n}{2}} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\hat{\sigma}} e^{-\frac{\hat{\varepsilon}_i^{*2}}{2\hat{\sigma}^2}} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\frac{\hat{\sigma}}{\sqrt{1-2\frac{\theta}{p}}}} e^{-\frac{\hat{\varepsilon}_i^{*2}}{2\frac{\hat{\sigma}^2}{1-2\frac{\theta}{p}}}}.$$

Recall that $\bar{Q}_\theta = Q_{-\theta}$. Then, the optimal titling point of our proposed method can be obtained by solving

$$E_{\bar{Q}_\theta}[T^* | T^* > a] = \psi_a'(\theta) = -a + \log(1 - 2\frac{\theta}{p})^{-\frac{n}{2}}. \qquad (4.21)$$

Instead of doing simulation under the normal family, we propose a transformed likelihood ratio method to obtain the bootstrap estimator of $P\{T > a\}$. The method is doing simulation under the χ^2-distribution family. Note that $T^* = \sum_{i=1}^{n} X_i^2/p$, where $X_i = \hat{\varepsilon}_i^{*2}/\hat{\sigma}^{*2}$ is a χ^2-distribution with degree of freedom 1. Here the likelihood ratio based on an exponential change of measure is

$$\frac{dQ}{dP} = e^{\theta(T^* - a) - \psi_a(\theta)} = \frac{e^{\theta(T^* - a)}}{Ee^{\theta(T^* - a)}}, \qquad (4.22)$$

where $\psi_a(\theta) = \log Ee^{\theta(T^* - a)}$. By using an argument similar as above, the optimal titling point can be obtained by solving

$$E_{\bar{Q}_\theta}[T^* | T^* > a] = \psi_a'(\theta) = -a + \log(1 - 2\frac{\theta}{p})^{-\frac{n}{2}}. \qquad (4.23)$$

Note that (4.23) equals (4.21).

Since the relative efficiencies of these two algorithms are almost the same, we only report that based on normal family. Table 5 reports the relative efficiency of the importance resampling with respect the naive Monte Carlo simulation. Here α denotes the true tail probability, a denotes the quantile of T^*, $\hat{\alpha}$ and $\widehat{var}_\alpha$ are the mean and variance of the probability estimator with Monte Carlo, $\hat{\alpha}(\theta)$ and $\widehat{var}_\alpha(\theta)$ are the mean and the variance of the tail probability estimator with importance sampling, $\text{RE}(\hat{\alpha}(\theta), \hat{\alpha})$ is the relative efficiency of $\hat{\theta}$ relative to $\hat{\alpha}$ in a regression model. Table 6 reports the non-coverage probabilities and averages and standard deviations of region areas for cubical and spherical confidence regions with nominal coverage probability 95%.

Table 5: Importance sampling with PMLE compared with naive method and a regression model in Monte Carlo simulation.

a	3.363	3.756	4.571
α	0.100	0.050	0.010
NV $\hat{\alpha}$	0.101	0.049	0.0099
$\widehat{var}_\alpha$	8.94E-04	4.74E-04	1.00E-04
PSD $\hat{\alpha}(\theta)$	0.099	0.050	0.010
$\widehat{var}_\alpha(\theta)$	1.98E-04	6.35E-05	3.76E-06
RE$(\widehat{\alpha}(\theta), \widehat{\alpha})$	4.51	13.39	26.59

Table 6: Non-coverage probabilities and averages and standard deviations of region areas for circular confidence regions with nominal coverage probability 95%. The sample is drawn from a regression model in Longley (1967). The parameter of interest is β. Results are reported for studentised statistics and naive resampling and importance resampling.

	Bootstrap		Region Area	
Resampling Method	replication size	Noncoverage probability	Average	Standard deviation
Naive	1000	0.054	49214.4	4853.05
Importance	400	0.049	51054.2	5035.12
Importance	200	0.051	52563.7	5174.63
Importance	100	0.053	54853.3	5313.38

5   Conclusions

In this paper, we propose a general account in importance sampling with applications to portfolio VaR computation and bootstrapping confidence regions. It is shown that our method produces efficient approximation to the problem. Simulation results confirm the theoretical results that our method always provides greater variance reduction than the naive Monte Carlo method. Our numerical experiments demonstrate that the gain in variance reduction can be substantial in some cases.

The key features of our method are twofold. First, (2.8) characterizes the optimal alternative distribution for importance sampling under exponential tilting. And the recursive algorithm facilitates the computation of the optimal solution. The initial value of the recursive algorithm is the dominating point of the large deviations tilting probability used previously by other authors; e.g., Sadowsky and Bucklew (1900). The recursive algorithm then sequentially generates alternative distributions providing greater variance reduction. Due to the nature of the recursive algorithm, the additional programming effort and computing time are negligible. Second, the

proposed tilting formula for normal distribution and its square, along with jump diffusion model, can be used to have more efficient simulation for portfolio VaR computation. We also apply the proposed titling probability to a parametric family (normal family or χ^2 family) which can be used for constructing bootstrap confidence regions for the unknown parameters in regression models.

Our method highlights the two aforementioned key features in general settings. By using the idea of conjugate probability measure, we obtained the optimal tilting parameter via (2.8). Specific considerations can be found in several papers; see Fuh and Hu (2004) for multivariate normal distributions, Fuh and Hu (2007) for hidden Markov models, and Fuh et al. (2011) for multivariate t distribution. Further applications to K-distributions and copula models will be published elsewhere. In this paper, we assume that the underlying random variables are independent over time. A more challenging project is to model the time dependence using, for example, Markov switching autoregression models or GARCH models. It is expected that our method can be applied to option pricing, Greeks letters calculation, and correlated default probabilities among others.

Appendix: Proof of Theorem 1

The existence of the optimization problem (2.3) can be proved by first showing that for $\theta \in \mathbf{R}$,

$$\psi'(\theta) \text{ is strictly increasing and } E_{\bar{Q}}[X|X \in A] \text{ is strictly decreasing.} \qquad (A.1)$$

To prove (A.1), we first note that $\psi(\theta)$ is the cumulant generating function of X, and therefore its second derivative $\psi''(\theta) > 0$ for $\theta \in \mathbf{I}$. This implies that $\psi'(\theta)$ is strictly increasing. Since $\Psi(\theta)$ is convex and steep by assumption, we have $\psi'(\theta) \to \infty$ as $\theta \to \theta_{\max}$. Furthermore, consider the conditional measure of $\bar{Q}$ on the set A, denoted by $\bar{Q}_A$, which is defined as

$$d\bar{Q}_A = \frac{\mathbf{1}_{\{X \in A\}} d\bar{Q}}{\int \mathbf{1}_{\{X \in A\}} d\bar{Q}}. \qquad (A.2)$$

Then we have

$$
\begin{aligned}
\frac{dE_{\bar{Q}_\theta}[X|X \in A]}{d\theta} &= \frac{d}{d\theta}\left(\frac{E[\mathbf{1}_{\{X \in A\}} X \mathrm{e}^{-\theta X}]}{E[\mathbf{1}_{\{X \in A\}} \mathrm{e}^{-\theta X}]}\right) \\
&= -\frac{E[\mathbf{1}_{\{X \in A\}} X^2 \mathrm{e}^{-\theta X}]}{E[\mathbf{1}_{\{X \in A\}} \mathrm{e}^{-\theta X}]} + \frac{E^2[\mathbf{1}_{\{X \in A\}} X \mathrm{e}^{-\theta X}]}{E^2[\mathbf{1}_{\{X \in A\}} \mathrm{e}^{-\theta X}]} = -\mathrm{var}_{\bar{Q}_A}(X) < 0.
\end{aligned}
$$

This implies that $E_{\bar{Q}_\theta}[X|X \in A]$ is strictly decreasing. The existence of the optimization problem (2.3) follows from $E_{\bar{Q}_0}[X|X \in A] = E[X|X \in A] > \mu = \psi'(0)$.

To prove the uniqueness, we note that the second derivative of G equals

$$
\begin{aligned}
\frac{\partial^2 G(\theta)}{\partial \theta^2} &= \frac{\partial^2}{\partial \theta^2} E[\mathbf{1}_{\{X \in A\}} \frac{dP}{dQ}] = \frac{\partial^2}{\partial \theta^2} E[\mathbf{1}_{\{X \in A\}} e^{-\theta X + \psi(\theta)}] \\
&= \frac{\partial}{\partial \theta} E[\mathbf{1}_{\{X \in A\}}(+\psi'(\theta)) e^{-\theta X + \psi(\theta)}] \\
&= E[\mathbf{1}_{\{X \in A\}}((-X + \psi'(\theta))^2 + \psi''(\theta)) e^{-\theta X + \psi(\theta)}]. \qquad (A.3)
\end{aligned}
$$

Since $\psi(\theta)$ is the cumulant generating function of X, its second derivative $\psi''(\theta) > 0$. It then follows from (A.3) that $\frac{\partial^2 G(\theta)}{\partial \theta^2} > 0$, which implies that there exists a unique minimum of $G(\theta)$.

To prove (2.8), we need to simplify the RHS of (2.5) under $\bar{Q}$. Standard algebra gives

$$\frac{E[\mathbf{1}_{\{X \in A\}} X e^{-\theta X}]}{E[\mathbf{1}_{\{X \in A\}} e^{-\theta X}]} = \frac{\int \mathbf{1}_{\{x \in A\}} x e^{-\theta x} dP / \tilde{\Psi}(\theta)}{\int \mathbf{1}_{\{x \in A\}} e^{-\theta x} dP / \tilde{\Psi}(\theta)} = \frac{\int \mathbf{1}_{\{x \in A\}} x d\bar{Q}}{\int \mathbf{1}_{\{x \in A\}} d\bar{Q}}. \tag{A.4}$$

As a result, (A.4) equals

$$\int \mathbf{1}_{\{x \in A\}} x d\bar{Q}_A = E_{\bar{Q}_A}[X] = E_{\bar{Q}}[X | X \in A],$$

which implies the desired result.

References

Asmussen, S. and P. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis.* Springer-Verlag, New York.

Beran, R. 1987. Prepivoting to reduce level error of confidence sets. *Biometrika* **74** 457-468.

Beran, R. and P. W. Millar. 1986. Confidence sets for a multivariate distribution. *Annals of Statistics* **14** 431-443.

Bucklew, J. A. 2004. *Introduction to Rare Event Simulation.* Springer-Verlag, New York.

Davison, A. C. 1988. Discussion of papers by D. V. Hinkley and by T. J. DiCiccio & J. P. Romano. *Journal of Royal Statistical Society* B **50** 356–357.

Do, K. A. and P. Hall. 1991. On importance resampling for the bootstrap. *Biometrika* **78** 161–167.

Duffie, D. and J. Singleton. 2003. *Credit Risk*, Princeton University Press.

Fuh, C. D. and I. Hu. 2004. Efficient importance sampling for events of moderate deviations with applications. *Biometrika* **91** 471–490.

Fuh, C. D. and I. Hu. 2007. Estimation in hidden Markov models via efficient importance sampling. *Bernoulli* **13** 492–513.

Fuh, C. D., I. Hu, Y. H. Hsu and R. H. Wang. 2011. Efficient simulation of Value at Risk with heavy-tailed risk factors. *Operations Research* **59** 1395-1406.

Glasserman, P., P. Heidelberger and P. Shahabuddin. 2000. Variance reduction techniques for estimating Value-at-Risk. *Management Science* **46** 1349–1364.

Glasserman, P., P. Heidelberger and P. Shahabuddin. 2002. Portfolio Value-at-Risk with heavy-tailed risk factors. *Mathematical Finance* **12** 239–269.

Hall, P. 1987. On the bootstrap and likelihood-based confidence regions. *Biometrika* **74** 481-493.

Hall, P. 1992. *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.

Heidelberger, P. 1995. Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Modeling and Computer Simulation* **5** 43–85.

Johns, M. V. 1988. Importance sampling for bootstrap conference intervals. *Journal of the American Statistical Association* **83** 709–714.

Jorion, P. 2001. *Value-at-Risk*, McGraw-Hill, New York.

L'Ecuyer, P., M. Mandjes and B. Tuffin. 2009. Editors. *Rare Event Simulation using Monte Carlo Mathods*, Rubino, G. and Tuffin, B. Editors. John Wiley & Sons, Ltd. New York.

Liu, J. 2001. *Monte Carlo Strategies in Scientific Computing,* Springer-Verlag, New York.

Longley, J. W. 1967. An appraisal of least squares programs for the electronic computer from the viewpoint of the user. *Journal of the American Statistical Association* **62** 819–841.

Neddermeyer, J. C. 2009. Computationally efficient nonparametric importance sampling. *Journal of the American Statistical Association* **104** 788–802.

Ney, P. 1983. Dominating points and the asymptotics of large deviations for random walk on $\mathcal{R}^d$. *Annals of Probability* **11** 158–167.

Sadowsky, J. and J. A. Bucklew. 1990. On large deviations theory and asymptotically efficient Monte Carlo simulation. *IEEE Transition of Information Theory* **36** 579–588.

Siegmund, D. 1976. Importance sampling in the Monte Carlo study of sequential tests. *Annals of Statistics* **4** 673–684.

Zhang, P. 1996. Nonparametric importance sampling. *Journal of the American Statistical Association* **91** 1245-1253.